arXiv:math/9911076v2 [math.CO] 7 Dec 1999

GROWTH IN FREE GROUPS (AND OTHER STORIES)

IGOR RIVIN

ABSTRACT. We start by studying the distribution of (cyclically reduced) elements of the free groups F_n with respect to their abelianization (or equivalently, their class in $H_1(F_n, \mathbf{Z})$). We derive an explicit generating function, and a limiting distribution, by means of certain results (of independent interest) on Chebyshev polynomials; we also prove that the reductions mod p (p – an arbitrary prime) of these classes are asymptotically equidistributed, and we study the deviation from equidistribution. We extend our techniques to a more general setting and use them to study the statistical properties of long cycles (and paths) on regular (directed and undirected) graphs. We return to the free group to study some growth functions of the number of conjugacy classes as a function of their cyclically reduced length.

INTRODUCTION

In this paper we begin by studying certain growth functions of the free group F_r, related to well-studied questions on the growth functions of geodesics on manifolds. The free group is a relatively simple combinatorial object, and this allows us to get fairly complete answers to our questions. Our techniques, which are quite elementary, allow us to get precise results on the distribution of elements in F_r as a function of their abelianization and in terms of their abelianization mod p. Our techniques turn out to be easily extensible to the study of paths in graphs with coefficients in compact groups.

Here is an outline of the paper: In Section 1 we set up an equivalence between counting cyclically reduced words on the free group F_r and counting circuits on an associated graph $\mathcal{G}_r$, which, in turn, involves understanding the spectrum of the adjacency matrix of $\mathcal{G}_r$ (of course the answer is easily obtained, and is well-known; for convenience we state it as Theorem 1.1). We use this framework to obtain a generating function for the number of elements of a fixed cyclically reduced length with prescribed abelianization (or homology class). This turns out to be essentially a Chebyshev polynomial of the first kind; see Definition 2.2 of the function R_r and Theorem 2.3 (a very brief introduction to Chebyshev polynomials is given in Section 3). The fact that the function $R_r(c; \mathbf{x})$ (at least for some special values of the parameter c) is a combinatorial generating function implies a previously unnoticed positivity result on Chebyshev polynomials; this result is generalized in

Section 4 in Theorems 4.1 and 4.2. Theorem 2.3 is used in Section 5 to derive a limiting distribution (as n tends to infinity) of cyclically reduced words length n among the possible homology classes. From the analytic standpoint this is also a qualitative result about Chebyshev polynomials, complementing the positivity Theorems 4.1 and 4.2. In Section 6 we show that if we study homology mod p, then the cyclically reduced words in F_r are asymptotically equidistributed among the p^r classes in $H_1(F_r, \mathbf{Z}/p\mathbf{Z})$. We also succeed in estimating the extent to which the cyclically reduced words in F_r are *not* equidistributed mod p (Section 6.1).

While the results in Sections 5 and 6 seem to depend on the explicit generating function that we have obtained, in Section 7 we show that our techniques are more general, and use them to study the equidistribution properties of long walks on regular graphs – we obtain a complete answer (Theorem 7.1) – and, without any change, closed orbits of irreducible primitive Markov processes (with a finite number of states). The arguments use elementary perturbation theory and the necessary technical results are contained in Section 10.

In Section 8 we extend our methods to study the functions defined on the *edges* of a graph, and as an application we derive the statistical properties of long walks *without backtracking* on the edges of an undirected graph.

We apply our methods to derive equidistribution results for long walks with coefficients in compact groups in Sections 7.1 and 9. Our results are completely explicit, in that knowing the irreducible representations of the group in question allows us to obtain complete asymptotics for the convergence to uniformity. Our results also apply, via the construction of a directed edge graph to the statistics of "geodesic", that is, backtrackless paths (Section 8). This, in turn, implies a result on the statistical properties of "primitive" orbits of Markov processes as above.

In Section 12 we point out real and philosophical applications of the above mentioned result to group theory (where this all started) and geometry.

Finally, in Sections 13-14.1 we derive a relationship between the number of cyclically reduced words and the number of conjugacy classes of bounded length. While the generating function of the first is a rational function, the generating function of the second is the integral of a Lambert series with an infinite number of poles. These results are then extended to a slightly more general case than that of free groups. We then (in Section 15) compute a zeta function for primitive conjugacy classes, and show that this *is* a rational function.

1. A MODEL AND A GENERATING FUNCTION

Let G be the free group $F_r = \langle a_1, \ldots, a_r \rangle$, and let $g \in G$ be an element. The defining property of G is that g is *uniquely* represented by a *reduced* word in $a_1, \ldots, a_r$, that is, a word where a_i is never adjacent to a_i^{-1} (Notation: in the sequel we shall write W for w^{-1}). We observe that such words over the alphabet $a_1, A_1, \ldots, a_n, A_n$ are, in turn, be generated by walks on the graph $\mathcal{G}_r$, constructed as follows: $\mathcal{G}_n$ has $2r$ vertices, labelled with the symbols $a_1, \ldots, a_r, A_r, \ldots, A_1$ – this peculiar order will simplify notation later. The vertex corresponding to a_i is connected by an edge to every vertex *except* A_i. In particular, there is a loop joining a_i to itself (so that $\mathcal{G}_r$ is not a *simple* graph). A walk $v_1 v_2 \ldots v_k$ gives the word $v_2 \ldots v_k$, so the correspondence between walks and words is a $2r - 1$-to-1 mapping. Note, however, that if we restrict our attention to closed walks (circuits with basepoint) on $\mathcal{G}_r$, then those are in bijective correspondence with *cyclically*

reduced words in G. In the sequel we will be interested exclusively with cyclically reduced words.

1.1. Counting cyclically reduced words.

To count cyclically reduced words, then, we need to count circuits in $\mathcal{G}_r$. This is a well-understood problem: If $\mathcal{A}_r$ is the adjacency matrix of $\mathcal{G}_r$, then the number of circuits of length k is equal to the trace of $\mathcal{A}_r^k$. To compute this trace we must compute the spectrum of $\mathcal{A}_r$, and to do this, it is better to write $\mathcal{A}_r = J_{2r} - P_r$, where J_N is an $N \times N$ matrix all of whose elements are 1 and P_r is the $2r \times 2r$ matrix such that

$$(P_r)_{ij} = \begin{cases} 1, & \text{if } i + j = 2r; \\ 0, & \text{otherwise.} \end{cases}$$

In order to compute the spectrum of $\mathcal{A}_r$, we note first that the matrix J_{2r} has rank 1. The kernel of J_{2r} is

$$\ker J_{2r} = \{(v_1, \dots, v_{2r}) \mid \sum_{i=1}^{2r} v_i = 0\},$$

while the vector $\mathbf{1} = (1, \dots, 1)$ is the eigenvector of eigenvalue $2r$.

The spectrum of P_r is not much more difficult to compute: The vector $\mathbf{1}$ is the eigenvector of P_r as well as of J_{2r}, this time with eigenvalue 1. To compute the rest of the spectral decomposition, let $\mathbf{x}$ be an eigenvector of P_r orthogonal to $\mathbf{1}$, and let λ be the corresponding eigenvalue. Then we have the following set of equations:

$$(1) \qquad\qquad \sum_{j=1}^{2r} x_j = 0$$

$$x_j = \lambda x_{2r-j+1}, \qquad j = 1, \dots 2r.$$

Since at least one of the x_j is not equal to zero, we see that $\lambda^2 = 1$, so $\lambda = \pm 1$. The orthogonality condition Eq. (1) can be rewritten as $\sum_{j=1}^{r}(1 + \lambda)x_j = 0$. Suppose $\lambda = -1$. Then, Eq. (1) holds *a forteriori*, and so the eigenspace of of -1 is r-dimensional. On the other hand, if $\lambda = 1$, then we have the additional constraint that $\sum_{j=1}^{r} x_j = 0$, so the eigenspace of 1 is $n - 1$ dimensional. Putting this all together, we see that the spectrum of the adjacency matrix $\mathcal{A}_r$ is $(2r - 1, \underbrace{1, \dots, 1}_{r}, \underbrace{-1, \dots, -1}_{r-1})$. We see therefore:

Theorem 1.1. *The number of cyclically reduced words of length m in F_r is equal to $(2r - 1)^m + 1 + (r - 1)[1 + (-1)^m]$.*

2. Counting cyclically reduced words in homology classes

Recall that the abelianization of F_r is $\mathbf{Z}^r$, generated by the classes of $[a_1], \dots, [a_r]$ of $a_1, \dots, a_r$ respectively. To compute the homology class of a word w in F_r we simply count the total exponents $e_1(w), \dots, e_r(w)$ of the generators used to write w. Then, $[w] = e_1(w)[a_1] + \cdots + e_r(w)[a_r]$. In this section we will compute the following generating function:

$$\mathcal{H}_r^{(k)}(x_1, \dots, x_r) = \sum_{w \in W_k} \prod_{i=1}^{r} x_i^{e_i(w)},$$

where the sum is taken over the set W_k of all cyclically reduced words w in $a_1, \dots, a_r, A_1, \dots, A_r$ of length k.

To compute $\mathcal{H}_r^{(k)}$, we return to circuits in $\mathcal{G}_r$. Given a circuit $c = v_1, \dots, v_k, v_{k+1} = v_1$, the contribution of c to $\mathcal{H}_r^{(k)}$ is the monomial m_c given by the following iterative procedure: we start with 1, every time we see the vertex a_i, we multiply m_c by x_i, and every time we see A_i, we multiply m_c by $1/x_i$. From this, it follows that:

Theorem 2.1. *The Laurent polynomial $\mathcal{H}_r^{(k)}$ is given by* $\operatorname{tr} B_r^k$, *where $B_r = D_r \mathcal{A}_r$, where, in turn,*

$$D_r = \begin{pmatrix} x_1 & & & & & & \\ & \ddots & & & & & \\ & & x_n & & & & \\ & & & 1/x_n & & & \\ & & & & \ddots & & \\ & & & & & 1/x_1 \end{pmatrix}$$

Computing the trace of B_r^k seems daunting at first, but one can use the approach we have used to prove Theorem 1.1.

First, note that
$$B_r = D_r \mathcal{A}_r = D_r J_{2r} - D_r P_r.$$
Evidently, the rank of $D_r J_{2r}$ is still equal to 1, and

$$\ker D_r J_{2r} = \{\mathbf{v} = (v_1, \dots, v_{2r}) \mid \sum_{j=1}^{2r} v_j = 0\}.$$

Note further that an eigenvector $\mathbf{v}$ of $D_r P_r$, such that $\mathbf{v} \in \ker D_r J_{2r}$, with associated eigenvalue λ, is also an eigenvector of B_r, with associated eigenvalue $-\lambda$. To find such an eigenvector, we must solve the system of equations:

$$\sum_{j=1}^{2r} v_j = 0$$
$$\lambda v_j = v_{2r-j+1}/x_j, \qquad j \leq r$$
$$\lambda v_j = v_{2r-j+1} x_j, \qquad j > r.$$

We find, as before, that $\lambda = \pm 1$. The first equation reduces to (almost as before) to

$$\sum_{j=1}^{r} v_j(1 + \lambda x_j) = 0,$$

so that the eigenspaces of both 1 and -1 are $(r-1)$-dimensional. What are the two remaining eigenvalues μ_1 and μ_2 of B_r? Note that since $\det D_r = 1$, we know that $\det B_r = \det \mathcal{A}_r$. Note now that $\det B_r = \mu_1 \mu_2 (-1)^{r-1}$, while $\det \mathcal{A}_r = (2r-1)(-1)^{r-1}$. So

$$(2) \qquad\qquad \mu_1 \mu_2 = 2r - 1.$$

On the other hand,

$$(3) \qquad\qquad \mu_1 + \mu_2 = \operatorname{tr} B_r = \sum_{j=1}^{r}(x_j + \frac{1}{x_j}).$$

Denoting $y_r = \frac{1}{2}\sum_{j=1}^{n}(x_j + 1/x_j)$, we see that μ_1, μ_2 are the two roots of the equation $z^2 - 2y_r z + (2r-1) = 0$, so that:

$$\begin{aligned} \mu_1 &= y_r - \sqrt{y_r^2 - (2r-1)}, \\ \mu_2 &= y_r + \sqrt{y_r^2 - (2r-1)}. \end{aligned}$$

The trace of B_r^k is then equal to $\mu_1^k + \mu_2^k + (r-1)[1+(-1)^k]$. This can be expressed in terms of well known special functions, if we make the substitution $y_r = \sqrt{2r-1}\,y_r'$. Then,

$$\mu_1^k = (2r-1)^{k/2}\left(y_r' - \sqrt{y_r'^2 - 1}\right)^k,$$

$$\mu_2^k = (2r-1)^{k/2}\left(y_r' + \sqrt{y_r'^2 - 1}\right)^k,$$

and so

$$\mu_1^k + \mu_2^k = (2r-1)^{k/2}\left\{\left(y_r' - \sqrt{y_r'^2 - 1}\right)^k + \left(y_r' + \sqrt{y_r'^2 - 1}\right)^k\right\}$$

$$= 2(\sqrt{2r-1})^k T_k(y_r'),$$

where $T_k(x)$ is the k-th Chebyshev polynomial of the first kind. To simplify notation in the sequel, we define:

Definition 2.2.

$$\begin{aligned} R_n(c; x_1, \ldots, x_k) &= T_n\left(\frac{c}{2k}\sum_{i=1}^{k}\left(x_i + \frac{1}{x_i}\right)\right) \\ S_n(c; x_1, \ldots, x_k) &= U_n\left(\frac{c}{2k}\sum_{i=1}^{k}\left(x_i + \frac{1}{x_i}\right)\right). \end{aligned}$$

And to summarize:

Theorem 2.3. *The number of cyclically reduced words of length k in F_r homologous to $e_1[a_1] + \cdots + e_r[a_r]$ is equal to the coefficient of $x_1^{e_1} \cdots x_r^{e_r}$ in*

(4) $$2\left(\sqrt{2r-1}\right)^k R_k\Big(\frac{r}{\sqrt{2r-1}}; x_1, \ldots, x_r\Big) + (r-1)[1+(-1)^k]$$

Remark. The rescaled Chebyshev polynomial $T_k(ax)/a^k$ is called the k-th Dickson polynomial $T_k(x, a)$ (see [Schur73]).

3. Some facts about Chebyshev polynomials

The literature on Chebyshev polynomials is enormous; [Rivlin90] is a good to start. Here, we shall supply the barest essentials in an effort to keep this paper self-contained.

There are a number of ways to define Chebyshev polynomials (almost as many as there are of spelling their inventor's name). A standard definition of the *Chebyshev polynomial of the first kind $T_n(x)$* is:

(5) $$T_n(x) = \cos n \arccos x.$$

In particular, $T_0(x) = 1$, $T_1(x) = x$. Using the identity

(6) $$\cos(x+y) + \cos(x-y) = 2\cos x \cos y$$

we immediately find the three-term recurrence for Chebyshev polynomials:

$$(7) \qquad\qquad T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x).$$

The definition of Eq. (5) can be used to give a "closed form" used in Section 2:

$$(8) \qquad\qquad T_n(x) = \frac{1}{2}\left[\left(x - \sqrt{x^2 - 1}\right)^n + \left(x + \sqrt{x^2 - 1}\right)^n\right].$$

Indeed, let $x = \cos\theta$. then $\left(x - \sqrt{x^2-1}\right)^n = \exp(-in\theta)$, while $\left(x + \sqrt{x^2-1}\right)^n = \exp(in\theta)$, so $\frac{1}{2}\left(x - \sqrt{x^2-1}\right)^n + \left(x + \sqrt{x^2-1}\right)^n = \Re\exp(in\theta) = \cos n\theta$.

Though we will not have too many occasions to use them, we also define Chebyshev polynomials of the second kind $U_n(x)$, which can again be defined in a number of ways, one of which is:

$$(9) \qquad\qquad U_n(x) = \frac{1}{n+1}T'_{n+1}(x).$$

A simple manipulation shows that if we set $x = \cos\theta$, as before, then

$$(10) \qquad\qquad U_n(x) = \frac{\sin(n+1)\theta}{\sin\theta}.$$

In some ways, Schur's notation $\mathcal{U}_n = U_{n-1}$ is preferable. In any case, we have $U_0(x) = 1$, $U_1(x) = 2x$, and otherwise the U_n satisfy the same recurrence as the T_n, to wit,

$$(11) \qquad\qquad U_{n+1}(x) = 2xU_n(x) - U_{n-1}(x).$$

From the recurrences, it is clear that for $f = T, U$, $f_n(-x) = (-1)^n f(x)$, or, in other words, every second coefficient of $T_n(x)$ and $U_n(x)$ vanishes. The remaining coefficients alternate in sign; here is the explicit formula for the coefficient $c^{(n)}_{n-2m}$ of x^{n-2m} of $T_n(x)$:

$$(12) \qquad c^{(n)}_{n-2m} = (-1)^m \frac{n}{n-m}\binom{n-m}{m}2^{n-2m-1}, \qquad m = 0, 1, \ldots, \left[\frac{n}{2}\right].$$

This can be proved easily using Eq. (7).

4. Analysis of the functions R_n and S_n.

In view of the alternation of the coefficients, the appearance of the Chebyshev polynomials as generating functions in Section 2 seems a bit surprising, since combinatorial generating functions have non-negative coefficients. Below we state and prove a generalization. Remarkably, Theorems 4.1 and 4.2 do not seem to have been previously noted.

Theorem 4.1. *Let $c > 1$. Then all the coefficients of $R_n(c; x)$ are non-negative. Indeed the coefficients of $x^n, x^{n-2}, \ldots, x^{-n+2}, x^{-n}$ are positive, while the other coefficients are zero. The same is true of S_n in place of R_n.*

Proof. Let a_n^k be the coefficient of x^k in $U_n((c/2)(x + 1/x))$. The recurrence gives the following recurrence for the a_n^k :

$$(13) \qquad\qquad a_{n+1}^k = c(a_n^{k-1} + a_n^{k+1}) - a_{n-1}^k.$$

Now we shall show that the following always holds:

(a): $a_n^k \geq 0$ (inequality being strict if and only if $n - k$ is even).
(b): $a_n^k \geq \max(a_{n-1}^{k-1}, a_{n-1}^{k+1})$, the inequality strict, again, if and only if $n - k$ is even.

(c): $a_n^k \geq a_{n-2}^k$ (strictness as above).

The proof proceeds routinely by induction; first the induction step (we assume throughout that $n - k$ is even; all the quantities involved are obviously 0 otherwise):

By induction $a_{n-1}^k < \min(a_n^{k-1}, a_n^{k+1})$, so by the recurrence 13 it follows that $a_{n+1}^k > \max(a_n^{k-1}, a_n^{k+1})$. (a) and (c) follow immediately.

For the base case, we note that $a_0^0 = 1$, while $a_1^1 = a_1^{-1} = c > 1$, and so the result for U_n follows. Notice that the above proof does *not* work for T_n, since the base case fails. Indeed, if b_n^k is the coefficient of x^k in $T_n((c/2)(x + 1/x))$, then $b_0^0 = 1$, while $b_1^1 = c/2$, not necessarily bigger than one. However, we can use the result for U_n, together with the observation (which follows easily from the addition formula for sin) that

$$(14) \qquad T_n(x) = \frac{U_n(x) - U_{n-2}(x)}{2}.$$

Eq. (14) implies that $b_n^k = a_n^k - a_n^{k-2} > 0$, by (c) above. $\qquad \square$

The proof above goes through almost verbatim to show:

Theorem 4.2. *Let $c > 1$. Then all the coefficients of R_n are non-negative. The same is true of S_n in place of R_n*

To complete the picture, we note that:

Theorem 4.3.

$$R_n(1; x) = \frac{1}{2}\left(x^n + \frac{1}{x^n}\right).$$

Proof. Let $x = \exp i\theta$. Then $1/2(x + 1/x) = \cos\theta$, and $R_n(1; x) = T_n(1/2(x + 1/x)) = \cos n\theta = 1/2(x^n + 1/x^n)$. $\qquad \square$

Remark 4.4. *For $c < -1$ it is true that all the coefficients of $R_n(c; .)$ and $S_n(c; .)$ have the same sign, but the sign is $(-1)^n$. For $|c| < 1$, the result is completely false. For c imaginary, the result is true. I am not sure what happens for general complex c.*

By the formula (12), we can write

$$(15) \qquad T_n\left(\frac{c}{2}\left(x + \frac{1}{x}\right)\right) = \frac{1}{2}\sum_{m=0}^{\left[\frac{n}{2}\right]}(-1)^m\frac{n}{n-m}\binom{n-m}{m}c^{n-2m}\left(x + \frac{1}{x}\right)^{n-2m}.$$

Noting that

$$(16) \qquad \left(x + \frac{1}{x}\right)^k = \sum_{i=0}^{k}\binom{k}{i}x^{k-2i}$$

we obtain the expansion

$$(17) \qquad R_n(c; x) = c^n\sum_{k=-n}^{n}x^k\sum_{m=0}^{\left[\frac{n}{2}\right]}\left(-\frac{1}{c^2}\right)^m\frac{n}{n-m}\binom{n-m}{m}\binom{n-2m}{(n-2m-k)/2},$$

where it is understood that $\binom{a}{b}$ is 0 if $b < 0$, or $b > a$, or $b \notin \mathbf{Z}$. We shall denote the coefficient of x^k by $t(n, k, c)$.

5. Limiting distribution of coefficients

While the formula (17) is completely explicit, and a similar (though somewhat more cumbersome) expression could be obtained for $R_n(c; x_1, \ldots, x_k)$, for many purposes it is more useful to have a limiting distribution formula as given by Theorem 5.1 below. To set up the framework, we note that since all the coefficients of $R_n(c; x_1, \ldots, x_k)$ are non-negative (according to Theorem 4.2), they can be thought of defining a probability distribution on the integer lattice $\mathbf{Z}^k$, defined by $p(l_1, \ldots, l_k) = [x_1^{l_1} x_2^{l_2} \cdots x_k^{l_k}] R(c; x_1, \ldots, x_k) / R(c; 1, \ldots, 1)$ (where the square brackets mean that we are extracting the coefficients of the bracketed monomial). Call the resulting probability distribution $\mathcal{P}_n(c; \mathbf{z})$, where $\mathbf{z}$ now denotes a k-dimensional vector.

Theorem 5.1. *With notation as above, when $c > 1$, the probability distributions $\mathcal{P}_n(c; \mathbf{z}/\sqrt{n})$ converge to a normal distribution on $\mathbf{R}^k$, whose mean is $\mathbf{0}$, and whose covariance matrix C is diagonal, with entries*

$$\sigma^2 = \frac{c}{k} \left[1 + \left(\frac{c+1}{c-1} \right)^{1/2} \right].$$

To prove Theorem 5.1 we will use the method of characteristic functions (Fourier transforms), and more specifically at first the *Continuity Theorem* ([FellerII, Chapter XV.3, Theorem 2]),

Theorem 5.2. *In order that a sequence $\{F_n\}$ of probability distributions converges properly to a probability distribution F, it is necessary and sufficient that the sequence $\{\phi_n\}$ of their characteristic functions converges pointwise to a limit ϕ, and that ϕ is continuous in some neighborhood of the origin.*

In this case ϕ is the characteristic function of F. (Hence ϕ is continuous everywhere and the convergence $\phi_n \to \phi$ is uniform on compact sets).

The characteristic function ϕ_n of $\mathcal{P}_n(c; \mathbf{z})$ is simply

$$R_n(c; \exp(i\theta_1), \ldots, \exp(i\theta_k)) / R(c; 1, \ldots, 1).$$

By definition of R_n,

$$R_n(c; \exp(i\theta_1), \ldots, \exp(i\theta_k)) \qquad = T_n \left(\tfrac{c}{k} \sum_{j=1}^k \cos \theta_j \right),$$

$$R_n(c; 1, \ldots, 1)) \quad = T_n \left(\tfrac{c}{k} \sum_{j=1}^k \cos 0 \right) = T_n(c).$$

We now use the form of Eq. (8):

$$T_n(x) = \frac{1}{2} \left(\left(x - \sqrt{x^2 - 1} \right)^n + \left(x + \sqrt{x^2 - 1} \right)^n \right),$$

setting

$$u = \sum_{j=1}^k \cos \frac{\theta_j}{\sqrt{n}},$$

and

$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k),$$

we get

$$(18) \quad \phi_n(\boldsymbol{\theta}/\sqrt{n}) = \frac{1}{T_n(c)} \left\{ \frac{1}{2}\left(\frac{c}{k}u + \sqrt{\frac{c^2}{k^2}u^2 - 1} \right)^n + \frac{1}{2}\left(\frac{c}{k}u - \sqrt{\frac{c^2}{k^2}u^2 - 1} \right)^n \right\}.$$

Notice, however, that for $c > 1$, the ratio of the second term in braces to the first is exponentially small as $n \to \infty$, since the first term grows like $(c + \sqrt{c^2 - 1})^n$, while the second as $(c - \sqrt{c^2 - 1})^n$ (since $\cos\frac{\theta_j}{\sqrt{n}} \to 1$). Since, for the same reason, $2T_n(c) = (c + \sqrt{c^2 - 1})^n[1 + o(1)]$, we can write:

$$\phi_n(\frac{\boldsymbol{\theta}}{\sqrt{n}}) = \left[\frac{\frac{c}{k}u + \sqrt{\frac{c^2}{k^2}u^2 - 1}}{c + \sqrt{c^2 - 1}} \right]^n + o(1).$$

Substituting the Taylor expansions for the cosine terms (hidden in u for typesetting reasons), we get:

$$(19) \qquad\qquad u = k + \frac{1}{2n}\langle \boldsymbol{\theta}, \boldsymbol{\theta} \rangle + o(1/n),$$

so

$$(20) \qquad\qquad \frac{c}{k}u = c + \frac{c}{2kn}\langle \boldsymbol{\theta}, \boldsymbol{\theta} \rangle + o(1/n).$$

A similar computation gives

$$(21) \qquad\qquad \frac{c^2}{k^2}u^2 = c^2 + \frac{c^2}{kn}\langle \boldsymbol{\theta}, \boldsymbol{\theta} \rangle + o(1/n).$$

Substituting the last expansion into the square root, we see that

$$\sqrt{\frac{c^2}{k^2}u^2 - 1} = \sqrt{c^2 - 1}\sqrt{1 + \frac{1}{n}\left[\frac{c^2}{(c^2-1)k}\langle \boldsymbol{\theta}, \boldsymbol{\theta} \rangle + o(\tfrac{1}{n}) \right]}$$
$$= \sqrt{c^2 - 1}\left[1 + \frac{1}{2n}\frac{c^2}{(c^2-1)k}\langle \boldsymbol{\theta}, \boldsymbol{\theta} \rangle \right] + o(\tfrac{1}{n}).$$

Adding Eq. (20) and collecting terms, get

$$(22)$$
$$\frac{\frac{c}{k}u + \sqrt{\frac{c^2}{k^2}u^2 - 1}}{c + \sqrt{c^2 - 1}} = 1 + \frac{1}{2n}\left(1 + \frac{1}{c + \sqrt{c^2 - 1}} \right)\left(\frac{c}{k} + \frac{c^2}{(c^2 - 1)^{1/2}k} \right)\langle \boldsymbol{\theta}, \boldsymbol{\theta} \rangle + o(\tfrac{1}{n}).$$

Performing some further simplifications, we see that

$$\phi_n(\frac{\boldsymbol{\theta}}{\sqrt{n}}) = \exp\left(-\frac{1}{2}\boldsymbol{\theta}^\perp C\boldsymbol{\theta} \right) + o(1),$$

where C is the covariance matrix described in the statement of Theorem 5.1, and Theorem 5.1 follows immediately.

Remark 5.3. The speed of convergence in Theorem 5.1 can be estimated using standard technology (see [FellerII, Chapter XVI], [Shiryaev96, Chapter III.11]), but the speed of convergence in practice (as checked by numerical experiments) seems to be much better than the general estimates. Indeed the L^1 difference between $\mathcal{P}_n$ and the normal distribution appears to decrease almost exactly linearly in n.

6. DISTRIBUTION MOD p

The explicit generating functions derived above can be used to study the distribution of cyclically reduced words in F_r with respect to their mod p-homology class (this is the analogue, in this setting, of the work of [PS87]).

Theorem 6.1. *Let h_1 and h_2 be two elements of $H_1(F_r, \mathbf{Z}/p\mathbf{Z}) = \mathbf{Z}/p\mathbf{Z}^r$, and let W_{r,n,h_1} and W_{r,n,h_2} be the numbers of cyclically reduced words in F_r homologous to h_1 and h_2, respectively. Then,*

$$(23) \qquad \lim_{n \to \infty} \frac{W_{r,n,h_2}}{W_{r,n,h_1}} = 1.$$

Proof. By elementary algebra (in one dimension, formula (25), the statement of theorem is equivalent to the statement that

$$(24) \qquad \lim_{n \to \infty} \frac{\phi_n(\boldsymbol{\theta})}{\phi_n(\mathbf{0})} = 0,$$

for $\boldsymbol{\theta} = (2n_1\pi/p, \ldots, 2n_r\pi/p)$, with not all n_j equal to 0 mod p, where ϕ_n is the characteristic function defined in the previous section.

The estimate of Eq. (24), however, follows immediately from the explicit formula (18): indeed, in the current context,

$$u(\boldsymbol{\theta}) = \sum_{j=1}^{k} \cos(2n_j\pi/p),$$

which is strictly smaller than $u(\mathbf{0})$, so the ratio of $\phi_n(\boldsymbol{\theta})$ to $\phi_n(\mathbf{0})$ goes to zero exponentially fast in n. $\qquad\square$

6.1. **Deviation from uniformity.** Although the distribution of homology mod p approaches uniformity, it turns out that there is a persistent *bias* in favor of certain homology classes. This is very much akin to the Chebyshev bias, analyzed in [RS94]. To simplify the discussion we project one more time: for each cyclically reduced word in F_r homologous to $a_1^{k_1} a_2^{k_2} \ldots a_r^{k_r}$ we consider $k_1 + \cdots + k_r$ mod p. In this case we have a univariate distribution, whose generating function is given by $\psi_n(x) = R_n(c; x, \ldots, x)$, with $c = \frac{r}{\sqrt{2r-1}}$ (as per formula (4; we leave in the general c, to underline that our results apply to general question on distribution of coefficients of the Laurent polynomials R_n).

The number of elements congruent to q mod p is given by

$$(25) \qquad \mathcal{N}_{n,q} = \frac{1}{p} \sum_{j=0}^{p-1} \chi^{-qj} \psi_n(\chi^j),$$

where $\chi = \exp(2\pi i/p)$ is a primitive p-th root of unity. Let us recall that

(26)
$$\psi_n(e^{ix}) = \frac{1}{T_n(c)} \left\{ \frac{1}{2} \left(c\cos x + \sqrt{c^2 \cos^2 x - 1} \right)^n + \frac{1}{2} \left(c\cos x - \sqrt{c^2 \cos^2 x - 1} \right)^n \right\}.$$

Note the following properties of the function ψ_n:

    (a): $\psi_n(1/x) = \psi_n(x)$,
    (b): If $c\cos x < 1$, then $|\psi_n(\exp(ix))|T_n(c) \le 1$.
    (c): $\psi_n\{\exp(i(\pi - x))\} = (-1)^n \psi_n\{\exp(ix)\}$
    (d): If $c\cos x \ge 1$, then $\psi_n(\exp(ix)) > 0$.

(e): If $x \in [0, \arccos 1/c]$, and n tends to infinity, then

$$(27) \qquad \frac{|\psi_n(\exp(ix))| T_n(c)}{\left[c + \sqrt{c^2 \cos^2 x - 1}\right]^n} = 1 + o(1),$$

and so

(f): $\psi_n(\exp(ix_1)) = o(\psi_n(\exp(ix_2)))$ for $0 \leq x_2 < \arccos 1/c$, and $x_2 < x_1 < \pi - x_2$.

Using property (a), we can write

$$(28) \qquad \mathcal{N}_{n,q} = \frac{1}{p} \left[\psi_n(1) + 2 \sum_{j=1}^{\frac{p-1}{2}} \cos \frac{2\pi q j}{p} \psi_n(\chi^j) \right].$$

Since $\cos \frac{2\pi m}{p} < 1$ is monotonically decreasing as a function of m for $0 \leq m \leq \frac{p-1}{2}$, we see:

Theorem 6.2. *For sufficiently large even n, $\mathcal{N}_{n,q} < \mathcal{N}_{n,0}$.*

Proof. This is an immediate consequence of the monotonicity of cos, equation (28) and properties (a), (c), (d), and (f) above. □

For $q \neq 0 \mod p$, the term largest in absolute value in the sum (aside the $\psi_n(1)$ term) on the right hand side of eq. (28) is the $\psi_{(\chi^{\frac{p-1}{2}})}$ term, so if we assume that n is even, then the next largest (after $\mathcal{N}_{n,0}$) term will be $\mathcal{N}_{n,p-2}$ (since $(p-2)[(p-1)/2] = 1 \mod p$), then $\mathcal{N}_{n,p-4}$, and so on. For n odd, the ordering is reversed.

7. An extension and limiting distributions for graphs

An inspection of the proof of Theorem 5.1 reveals that in order to show that for a sequence of probability distributions $\{P_n(x)\}$ on $\mathbf{Z}$, the distributions $\{P_n(x/\sqrt{n})\}$ converged to a limiting normal distribution with mean 0, we used the following conditions (we will state them in a univariate setting for simplicity; the multivariate case is the same):

Condition 1. The characteristic function of $\{P_n\}$ has the form

$$\chi(P_n) = f^n(\theta) + o(1),$$

where $f_j(\theta)$ is twice continuously differentiable at 0, so that $f_j(\theta) = a_j + b_j \theta + c_j \theta^2 + o(\theta^2)$.

Condition 2.

$$a_1 = 1, \qquad b_2 = 0, \qquad c_2 < 0.$$

Suppose now we generalize the setting of Section 1 as follows:

Let $\mathcal{G}$ be a connected r-regular non-bipartite graph, directed or not, (possibly with self-loops and multiple edges), on k vertices. Let v_1 and v_2 be two vertices of $\mathcal{G}$. Consider now the set W_N of all closed walks (circuits) of length N on $\mathcal{G}$. Let $\mathbf{f} : V(\mathcal{G}) \to R$ be a function assigning a weight to each vertex of $\mathcal{G}$, and define a random variable $X_{\mathbf{f}}$ to be $\sum_{l=1}^{N} \mathbf{f}(v_l)$ for $w = v_1, \dots, v_N \in W_N$. What can we say about the distribution of $X_{\mathbf{f}}$? It turns out that asymptotically we can say a lot. First, however, define

$$\mu(\mathbf{f}) = \frac{1}{k} \sum_{j=1}^{k} \mathbf{f}(v_j),$$

and $\mathbf{f}_0 = \mathbf{f} - \mu(\mathbf{f})\mathbf{1}$. Define further the *Laplacian* $\Delta(\mathcal{G})$ of $\mathcal{G}$ to be $\Delta(\mathcal{G}) = r\mathbf{I} - A(\mathcal{G})$, and define $\Delta_0(\mathcal{G})$ to be $\Delta(\mathcal{G})$ viewed as an operator on the orthogonal complement to $\mathbf{1}$ (that is, vectors with 0 sum). Let $P_N(x)$ be the distribution of $X_{\mathbf{f}}$ on W_N.

Theorem 7.1. *The distributions $P_N((x - N\mu(\mathbf{f}))/\sqrt{N})$ converge to a balanced (that is, mean 0) normal distribution with variance*

$$(29) \qquad \begin{aligned} \sigma^2(\mathbf{f}) &= \frac{1}{k}\left[-\|\mathbf{f}_0\|^2 + 2r\mathbf{f}_0^t\Delta_0^{-1}(\mathcal{G})\mathbf{f}_0\right] \\ &= \frac{1}{k}\left[\mathbf{f}_0^t(-\mathbf{I}_0 + 2r\Delta_0^{-1}(\mathcal{G}))\mathbf{f}_0\right]. \end{aligned}$$

Proof. Exactly as in Section 1 we construct a generating function g_N for $X_{\mathbf{f}}$ on W_N. To do this, let A be the adjacency matrix of $\mathcal{G}$, and let

$$D_k(x) = \begin{pmatrix} x^{\mathbf{f}(v_1)} & & & & \\ & x^{\mathbf{f}(v_2)} & & & \\ & & x^{\mathbf{f}(v_3)} & & \\ & & & \ddots & \\ & & & & x^{\mathbf{f}(v_k)} \end{pmatrix}.$$

Then, $g_N(x) = \operatorname{tr}(D_k(x)A)^N = \sum_{j=1}^k \lambda_j^N(D_k(x)A)$, where $\lambda_1, \ldots, \lambda_j$ are eigenvalues, and, just as in Section 5, we have $\chi(P_N)(\theta) = g_N(\exp(i\theta))/c_N$, where

$$c_N = |W_N| = \sum_{j=1}^k \lambda_j^N(A).$$

Since $\mathcal{G}$ is an *r=regular*, *non-bipartite* graph, it has a unique eigenvalue of maximal modulus, and that eigenvalue is $\lambda_1 = r$.

Now, we can directly apply Conditions 1 and 2 (and accompanying comments) above, and the results of Section 10 (noting that Assumptions 1–4 hold) to obtain the desired result (in particular, the estimate needed in Condition 2 is precisely Theorem 10.8). We replaced the resolvent in formula (54) by the equivalent (by the discussion in the beginning of Section 10) Laplacian form, since that is more common in graph theory. $\qquad\square$

Remark 7.2. *If the vector $\mathbf{f}$ is an eigenvector of A^tA with eigenvalue r^2, the corresponding variance is equal to zero. By Remark 10.9 this will not happen, eg, if G is a connected non-bipartite undirected graph, but it does happen for general directed graphs; see the discussion of the directed line graph in Section 8.*

The above remark leads to the following

Question: What combinatorial property of an r-regular directed graph G is reflected in the algebraic statement that the operator norm of $A_0(G)$ is equal to r?

A slight change in notation transforms Theorem 7.1 into a central limit theorem for distributions over closed orbits of primitive irreducible Markov processes over a finite number of states – the irreducibilty is exactly equivalent to the connectivity of the graph $\mathcal{G}$ above. For ease of reference we state this as a separate theorem. The notation for $\mathbf{f}$, μ, etc, is as before; the space W_N is now a probability space with the obvious probability measure; $\mathbf{P} = \mathbf{P}^t$ is the transition matrix (note that Remark 7.2 remains valid in this setting as well).

Theorem 7.3. *Let $P_N(x)$ be the distribution of $X_{\mathbf{f}}$ on W_N. Then $P_N((x-N\mu(\mathbf{f}))/\sqrt{N})$ converge to a balanced (that is, mean 0) normal distribution with variance*

$$(30) \quad \sigma^2(\mathbf{f}) = \frac{1}{k}\left[-\|\mathbf{f}_0\|^2 + 2\mathbf{f}_0^t(\mathbf{I}_0 - \mathbf{P}_0)^{-1}\mathbf{f}_0\right] = \frac{1}{k}\left[\mathbf{f}_0^t(-\mathbf{I}_0 + 2r(\mathbf{I}_0 - \mathbf{P}_0)^{-1})\mathbf{f}_0\right].$$

Remark 7.4. We have actually shown a slightly stronger result: instead of the trace (distribution over cycles), we could have considered the ij-th element of $\mathbf{P}$. Since the principal eigenvector varies continuously under perturbations (see [Kato66, Chapter II.4.1]), we could have replaced our sample space W_N as above by the space $\mathcal{C}_N$ of paths of length N joining the i-th to the j-th vertex. An easy computation shows that the covariance is the covariance of given in equation 30, divided by a further factor of k. The same remark applies to Theorem 7.1.

7.1. **Distribution modulo a prime.** Theorems 7.1 and 7.3 have particularly simple analogues if the function f we are studying is integer valued, and we are interested in the distribution of the $\mathbf{Z}/p\mathbf{Z}$-valued random variable $Y_f(n)$ which assigns to each cycle of length n the sum of the values of f modulo p. In that case, under the assumption that the adjacency matrix A (in the context of Theorem 7.1) or the transition matrix A (in the context of Theorem 7.3) is irreducible and primitive (the last two $A(\mathcal{L}_u(G))$ conditions guarantee that A has a single eigenvalue λ_0 of maximal modulus, the eigenspace of λ_0 is one-dimensional, and the orthogonal subspace is invariant under A), then we see that the distributions $\mathcal{P}_n$ of $Y_f(n)$ approach the uniform distribution (on $\mathbf{Z}/p\mathbf{Z}$) exponentially fast in n (though a more reasonable measure of the speed of convergence is the size of W_n, in which case the convergence is polynomial). This statement follows from the:

Lemma 7.5. *If A is a matrix satisfying the conditions above, then the spectral radius r_{UA} of UA, for U any non-trivial unitary matrix such that the top eigenvector of A is not also an eigenvector of U, is strictly smaller than that of A (r_A).*

The proof of the lemma is immediate.

In our case, the matrix U is the diagonal matrix $U(\chi)$ with $u_j j = \chi_p^{f_j}$, with χ_p a non-trivial p-th root of unity. The speed of convergence to the uniform distribution is given by $(\max_{\chi^p=1} r(U(\chi)A))/r(A)$.

8. FUNCTIONS ON EDGES AND DISTRIBUTIONS OVER PATHS WITHOUT BACKTRACKING

In this section we consider two kinds of questions, which are seen to be intimately related. The first is:

Question 1. Let f be a function on the *edges* of G. How are the averages of f over long cycles or paths in G distributed?

The second sort of question is:

Question 2. Let f be a function on the *vertices* of G. How are the averages of f distributed over long cycles in G *without backtracking* – such cycles are more closely related to, *eg*, geodesics on surfaces, then arbitrary cycles.

Both questions can be answered at the same time by constructing the *directed line graph* (or *line digraph*) of G. This construction can be performed for either a directed or undirected graph G; In section 8.1 we will derive the results for undirected graphs in detail, whilst in section 8.3 we will discuss the directed case somewhat more briefly (since the technical details are essentially identical).

8.1. **The directed line graph of an undirected graph.** The *directed line graph* of G, denoted by $\mathcal{L}(G)$, is constructed as follows: The vertices of $\mathcal{L}(G)$ are edges of G labelled with a $+$ or a $-$; that is, to each edge e of G there correspond vertices e_- and e_+ of $\mathcal{L}(G)$. These correspond to the two possible orientations of e: if the vertices of e are v and w, then we say that v is the head of e_-, and w the tail (and write $v = h(e_-)$, $w = t(e_-)$), while for e_+ this nomenclature is reversed. Two vertices v_1 and v_2 of $\mathcal{L}(G)$ are joined by a (directed) edge if the head of v_1 is the same as the tail of v_2, except that e_- is never joined to e_+, and *vice versa*. We now make some observations and definitions.

Definition 8.1. *Let f be a function defined on the vertices of a graph G. We say that a function g defined on the vertices of $\mathcal{L}(G)$ is the gradient of f, and write $g = \nabla f$ if $g(e) = f(h(e)) - f(t(e))$.*

Definition 8.2. *We can identify functions on the vertices of G with (a subset of) functions on the the vertices of $\mathcal{L}(G)$. To wit, if a f is a function on the vertices of G, we let $\mathcal{L}f(e) = f(h(e))$.*

Observation 8.3. *There is a natural correspondence between walks on $\mathcal{L}(G)$ and walks on G without backtracking. Indeed, passing through a vertex e of $\mathcal{L}(G)$ corresponds to going from $t(e)$ to $h(e)$. Since e_+ is not connected to e_- for any $e \in E(G)$, any such walk is automatically without backtracking. Similarly, a cycle on $\mathcal{L}(G)$ corresponds to a tailless cycle without backtracking on G.*

If G is an r-regular graph, then $\mathcal{L}(G)$ is $r-1$-regular, in the strong sense: each vertex of $\mathcal{L}(G)$ has in-degree and out-degree equal to $r-1$ (thus the total degree is $2r-2$), and from the above Observation 8.3, $\mathcal{L}(G)$ is connected if and only if G is. It follows that the adjacency matrix $A(\mathcal{L}(G))$ of $\mathcal{L}(G)$ is an irreducible nonnegative matrix, all of whose row and column sums are equal to $r-1$. It follows that the space of functions on the vertices of $\mathcal{L}(G)$ orthogonal to the vector $\mathbf{1}$ is an invariant subspace of $A(\mathcal{L}(G))$ and of $A^t(\mathcal{L}(G))$ – we will, as before, denote the two matrices restricted to this subspace by A_0 and A_0^t, respectively; the algebraic and geometric multiplicities of the eigenvalue $r-1$ is equal to 1, by standard Perron-Frobenius theory. Despite this, it turns out that $A^t A$ is spectacularly degenerate. Indeed, the ij-th entry of $A^t A$ is equal to the number of vertices of $\mathcal{L}(G)$ adjacent simultaneously to the i-th and the j-th vertex. It follows that the ii-th entry of $A^t A$ is equal to $r-1$, while the ij-th entry is equal to $r-2$ if the corresponding directed edges of G have the same tail, and is 0 otherwise. It follows that

$$
(31) \qquad A^t A = I_{2E(G)} + (r-2) \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_{V(G)} \end{pmatrix},
$$

where the last term contains $V(G)$ $r \times r$ blocks, each of which is the matrix of all 1s. We thus have the following observation:

Observation 8.4. *The spectrum of $A^t A$ has the following form: The eigenvalue $(r-1)^2$ occurs $V(G)$ times, and the corresponding eigenvectors are given precisely by $\mathcal{L}f$ for arbitrary functions f on G (the Perron eigenvector corresponding to the constant function), while the eigenvalue 1 occurs $2E(G) - V(G)$ times. The*

eigenvectors are those functions on the directed edges of G, for which, for all vertices v of G, the sum of values on all the edges leaving v is equal to 0.

Corollary 8.5. *The operator norm of A_0 is equal to $r - 1$.*

Consider now the Laplace operator on $\mathcal{L}(G)$: $\Delta_{\mathcal{L}(G)} = (r-1)I - A(\mathcal{L}(G))$. We will need the following in the sequel:

Theorem 8.6. *Let E_{r-1} be the eigenspace of $(r-1)^2$ for $A^t A$. If $V^*(G)$ is the space of functions on the vertices of G, then*
  (a):
$$E_{r-1} = \mathcal{L}(V^*(G)),$$
  (b):
$$\Delta_{\mathcal{L}(G)}(E_{r-1}) = \nabla(V^*(G)),$$
  (c): $\nabla(V^*(G)) \cap E_{r-1} \cap \mathbf{1}^\perp = \emptyset$, *unless G is bipartite.*

Proof. Part (a) is the content of Observation 8.4. Part (b) is a corollary of Part (a). Indeed, $\Delta_{\mathcal{L}(G)}(f)(x) = (r-1)f(x) - \sum_{h(x)=t(y)} f(y)$. If $f = \mathcal{L}g$, then

(32) $$\Delta_{\mathcal{L}(G)}(f)(x) = (r-1)(g(t(x)) - g(h(x))),$$

since all the y adjacent to x have the same tail, equal to the head of x.

To show Part(c), suppose $\nabla(V^*(G)) \cap E_{r-1} \neq \emptyset$. Let g be in the intersection, and k be such that $\nabla(k) = g$. It follows that for any x, y such that $t(x) = t(y)$, $g(x) = g(y)$. We see that $k(h(x)) - k(t(x)) = k(h(y)) - k(t(y))$, which implies in turn that $k(h(x)) = k(h(y))$. So, k is the eigenvector of the 0 eigenvalue of the Laplace operator on G, and hence is constant, unless G is bipartite. $\qed$

We end this section with a remark necessary to compute distributions, as done in the following Section 8.2. To wit:

Remark 8.7. *The adjacency matrix of the line graph of a non-bipartite graph G is primitive. That is, there is only one eigenvalue on the circle of radius $r - 1$ in the complex plane, and that is $r - 1$. Its geometric multiplicity is 1.*

Proof. Doubtlessly there are simpler arguments, but we choose to use the results (described in [ST96]) on the Ihara zeta function Z of G, which can be expressed as a determinant in two ways:

The first way (original theorem of Ihara [Ihara66]) is:

(33) $$Z^{-1}(u) = (1 - u^2)^{\mathcal{R}-1} \det((1 + (r-1)u^2)\mathbf{I} - uA),$$

with A the adjacency matrix of G, and $\mathcal{R}$ the rank of the fundamental group of G.

The second way (due to Hyman Bass [Bass92] is):

(34) $$Z^{-1}(u) = \det(\mathbf{I} - uM),$$

where M is the adjacency matrix of the directed line graph of G

The equality of the two expressions implies that v is an eigenvalue of M if and only if $v + (r-1)/v$ is an eigenvalue of A (we are ignoring the eigenvalues ± 1, which occur with large multiplicity in the spectrum of M). Suppose that v has modulus $r - 1$, so that $v = (r-1)\exp(i\theta)$, for some θ. It follows that

$w = \exp(i\theta) + (r - 1)\exp(-i\theta)$ is an eigenvalue of A, and since A is symmetric, $\theta \in \{0, \pi\}$. If $\theta = 0$, $v = r - 1$, while if $\theta = \pi$, $v = -(r - 1)$, but then $w = -r$ is an eigenvalue of A, and so G is bipartite.

The statement about the multiplicity of the eigenvalue $r - 1$ is immediate, since $\mathcal{L}(G)$ is clearly strongly connected. $\qquad\square$

We include the following observations both for the sake of completeness, and in view of Lemma 8.12 below.

Lemma 8.8.

$$\Delta\mathcal{L} = (r - 1)\nabla.$$

Proof. Indeed, $\mathcal{L}(f)(x) = f(t(x))$. Further,

(35)
$$\Delta\mathcal{L}(f)(x) = \sum_{t(y)=h(x)} f(t(x)) - f(t(y)) = (r-1)(f(t(x)) - f(h(x))) = \nabla(f)(x).$$

$\qquad\square$

Lemma 8.9. *For any $f, g \in V^*(G)$, we have*

$$(\mathcal{L}f)^t \nabla g = f^t \Delta g.$$

Proof. Indeed,

$$\begin{aligned}
(\mathcal{L}f)^t \nabla g &= \sum_x (f(t(x))(g(t(x)) - g(h(x)))) \\
&= \sum_{v \in V(G)} \sum_{w \text{ adjacent to } v} f(v)g(v) - f(v)g(w) \\
&= \sum_{v \in V(G)} f(v)\Delta(g)(v) \\
&= f^t \Delta g.
\end{aligned}$$

(36)

$\qquad\square$

Consider now a function g on the directed edges of G. How do we decompose it into a gradient and a function orthogonal to gradients? First, we note that a basis of the gradients is formed by the gradients of δ functions:

(37)
$$\delta_v(x) = \begin{cases} 1 & x = v, \\ 0 & \text{otherwise.} \end{cases}$$

So that

(38)
$$\nabla\delta_v(x) = \begin{cases} 1 & t(x) = v, \\ -1 & h(x) = v, \\ 0 & \text{otherwise.} \end{cases}$$

The functions $\nabla\delta_v$ form a basis of $\nabla(V^*(G))$, though not an orthonormal one. Now, note that

$$g^t \nabla\delta_v = \sum_{t(x)=v} g(x) - \sum_{h(y)=v} g(y).$$

In other words,

Lemma 8.10. *g is orthogonal to the gradients, if and only if the sum of g over the edges coming into any vertex v is equal to the sum of g over the edges leaving v. An equivalent condition is that $\nabla^t g = 0$.*

One may ask: what is the orthogonal projection of a given $\mathcal{L}f$ onto the gradients? The following comes out of an easy computation:

Observation 8.11. *The orthogonal projection of $\mathcal{L}f$ onto the set of gradients is $\nabla \Delta f$.*

8.2. **Applications to distribution.** We can use the results of the previous section to understand the limiting distribution of functions defined on (directed) edges of G. Indeed, we can use Theorem 7.1 in the form corresponding to Eq. 55 to observe that

$$(39) \qquad \sigma^2(\mathbf{f}) = \frac{1}{2rk}\mathbf{f}^t(\Delta_0^{-1})^t((r-1)^2\mathbf{I} - A(\mathcal{L}(G))^t A(\mathcal{L}(G))\Delta_0^{-1}\mathbf{f}$$

for $\mathbf{f}$ any function on the directed edges of G, and Δ_0 the restriction of the Laplace operator on $\mathcal{L}(G)$ to the subspace of 0-sum vectors.

Lemma 8.12. *The right hand sidef of equation 39 vanishes precisely when $\mathbf{f}$ is the gradient of a function on the vertices of G.*

Proof. Let $\mathbf{f} = \Delta u$. By Observation 8.4 we see that the right hand side of Eq. 39 vanishes precisely if $u \in \mathcal{L}(V^*(G))$. By part (b) of Theorem 8.6 it follows that this is so if and only if $\mathbf{f} \in \nabla(V^*(G))$. □

One direction of the above lemma is just common sense, since the sum over any cycle of a gradient is equal to 0.

Keeping the above in mind, we note that a simpler form of the covariance is given by Theorem 7.1:

$$(40) \qquad \sigma^2(\mathbf{f}) = \frac{r-1}{2rk}\left[\mathbf{f}^t\left(\mathbf{I} - 2(r-1)\Delta_0^{-1}\right)\mathbf{f}\right]$$

For functions on the vertices of G, the above assumes the form:

$$(41) \qquad \sigma^2(\mathbf{f}) = \frac{r-1}{2rk}\left[\mathbf{f}^t\mathcal{L}^t\left(\mathbf{I} - 2(r-1)\Delta_0^{-1}\right)\mathcal{L}\mathbf{f}\right]$$

8.3. **The line graph of a directed graph.** The construction of the line graph of a directed graph G is essentially the same as that of an undirected graph. This time, the vertices of $\mathcal{L}(G)$ without labels (so $\mathcal{L}(G)$ has $E(G)$ vertices). The operators ∇ and $\mathcal{L}$ are defined as in Section 8.1. We have an observation even simpler than Observation 8.3:

Observation 8.13. *There is a natural bijective correspondence between walks on $\mathcal{L}(G)$ and walks on G.*

If G is an r-regular directed graph (by this we mean that both the in- and out-degree of each vertex is equal to r), then so is $\mathcal{L}(G)$; by Observation 8.13 $\mathcal{L}(G)$ is connected whenever G is. As before, $A(\mathcal{L}(G))$ is the adjacency matrix of $\mathcal{L}(G)$. we can compute:

$$(42) \qquad A^t A = r \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_{V(G)} \end{pmatrix},$$

where each block corresponds to the set of edges of G emanating from a given vertex. From this we have:

Observation 8.14. *The spectrum of $A^t(\mathcal{L}(G))A(\mathcal{L}(G))$ has the following form: The eigenvalue r^2 occurs $V(G)$ times, and the corresponding eigenvectors are given by $\mathcal{L}f$ for arbitrary functions f on G (The Perron eigenvector corresponding to the constant function) while the eigenvalue 0 occurs $E(G) - V(G)$ times. The eigenvectors are those functions on the edges of G for which the sums of the values over all edges leaving a vertex v is equal to 0 (for all v).*

Corollary 8.15. *The operator norm of $A_0(\mathcal{L}(G))$ is equal to r.*

The Laplace operator on $\mathcal{L}(G)$ is defined as: $\Delta_{\mathcal{L}(G)} = rI - A(\mathcal{L}(G))$.
We have

Theorem 8.16. *Let E_r be the eigenspace of r^2 for $A^t A$. If $V^*(G)$ is the space of functions on the vertices of G, then*

(a):
$$E_r = \mathcal{L}(V^*(G)),$$

(b):
$$\Delta_{\mathcal{L}(G)}(E_r) = \nabla(V^*(G)),$$

We also include

Remark 8.17. *The adjacency matrix of the line graph of G is primitive if the adjacency matrix of G is.*

Proof. We use Observation 8.13 and Theorem 15.1 to note that the non-zero eigenvalues of G are exactly the same as those of $\mathcal{L}(G)$, since $\det(I - uA(G)) = \det(I - uA(\mathcal{L}(G)))$. $\qquad \square$

Lemma 8.18.
$$\Delta\mathcal{L} = r\nabla.$$

Lemma 8.19. *For any $f, g \in V^*(G)$, we have*
$$(\mathcal{L}f)^t \nabla g = f^t \Delta g.$$

Lemma 8.10 and Observation 8.11 go through without change.

The results of section 8.2 go through essentially without change. Since some constants change we restate them here. First, let f be a function defined on the edges of $\mathcal{L}(G)$. We see that:

$$(43) \qquad \sigma^2(\mathbf{f}) = \frac{1}{2rk}\mathbf{f}^t(\Delta_0^{-1})^t r^2 \mathbf{I} - A(\mathcal{L}(G))^t A(\mathcal{L}(G))\Delta_0^{-1}\mathbf{f}$$

Lemma 8.12 holds as well, and this gives us the following useful corollary (a homological condition) about distribution on G itself:

Theorem 8.20. *The variance of a function f on the vertices of G vanishes, precisely when there exists a function g, such that $\mathcal{L}f = \nabla g$.*

Finally, we have a version of formula 40:

$$\sigma^2(\mathbf{f}) = \frac{1}{2k} \left[\mathbf{f}^t \left(\mathbf{I} - 2r\Delta_0^{-1} \right) \mathbf{f} \right] \tag{44}$$

9. Distribution in compact groups

The methods of the section 7.1 can be adapted to the following setting: Let G is a graph, and T be a compact topological group. Label the i-th vertex of G with $t_i \in T$. Now, associate to each cycle $c = v_1, \ldots, v_k$ on G the element $t_c = t_k \cdot \ldots \cdot t_1 \in T$. We ask: as c varies over the cycle space W_N, how are the elements t_c distributed in T (with respect to the Haar measure). The answer is given by the following:

Theorem 9.1. *If the graph G is as before (connected, non-bipartite), the closed subgroup generated by the t_i $(i = 1, \ldots, k)$ is equal to G, and the elements t_i do not all lie in the same coset with respect to a one-dimensional representation of G, then the elements t_c become equidistributed, as $N \to \infty$.*

Proof. As before, the equidistribution is equivalent to the assertion that for a *nontrivial* irreducible unitary representation ρ,

$$\sum_{c \in W_n} \mathrm{tr}\left(\rho(t_c)\right) = o(|W_n|). \tag{45}$$

This follows from the Fourier transform formula for compact groups; see [FH91] for the finite case, [Weil51] for the general compact topological group case. Now, let $U(\rho)$ be the $k\deg\rho \times k\deg\rho$ block-diagonal matrix whose j-th block is just $\rho(t_j)$. Further more, as before, let $A(G)$ be the adjacency matrix of G, and $A_l(G) = A(G) \otimes \mathbf{I}_l$ (where $\mathbf{I}_l$ is the $l \times l$ diagonal matrix: in other words, $A_l(G)$ is a $kl \times kl$ matrix, obtained from $A(G)$ by replacing each element a_{ij} by a $k \times k$ matrix M_{ij}, all of whose elements are equal to a_{ij}. It is not hard to see that the left hand side of Eq. 45 is equal to $\mathrm{tr}\left(U(\rho)A_{\deg\rho}(G)\right)^N$, and so it suffices to show that the spectral radius of $M_\rho = U(\rho)A_{\deg\rho(G)}$ is strictly smaller than the spectral radius of $A(G)$ (which we normalize to be equal to 1 by scaling) under the hypotheses of the theorem. Suppose not. Since $(U(\rho))$ is unitary, the worst that can happen is that there exists a unit vector v, such that $\|M_\rho(v)\| = 1$. If that is so, v is contained in the eigenspace of eigenvalue 1 of $A_{\deg\rho}$. In such a case, $v = v_1 \otimes u$, where $u \in V(\rho)$, and v_1 is an eigenvector of $A(G)$ with eigenvalue 1. If $v_1 = (v_1^1, \ldots, v_1^n)$, then $v_1^i u$ must be an eigenvector of $\rho(t_i)$, for all i. Since $v_1^i \neq 0 \, \forall i$, this implies that u is an eigenvector $\rho(t_i)$, $\forall i$. Since ρ is irreducible, this implies that *either* the elements $t_1, \ldots, t_k$ do not generate all of T, or ρ is 1-dimensional, in which case clearly $\rho(t_i) = \rho(t_j)$, $\forall i, j$, which proves the theorem. $\square$

Remark 9.2. As in Remark 7.4, the above argument also works if we pick all paths between the i-th and the j-th vertex of G, instead of all cycles.

10. SOME PERTURBATIONS AND ESTIMATES

Consider an analytic family of linear operators $M(x)$, acting on $\mathbf{R}^k$, with $M(0) = M$, and let λ be a simple eigenvalue of M. Then, if

$$M(x) = M + M^{(1)}x + M^{(2)}x^2 + \dots,$$

perturbation theory (see [Kato66, page 79, (2.33)]) tells us that

$$\lambda(x) = \lambda + \lambda^{(1)}x + \lambda^{(2)}x^2 + \dots,$$

where

$$(46) \qquad \lambda^{(1)} \qquad = \operatorname{tr} M^{(1)}P_\lambda,$$

$$(47) \qquad \lambda^{(2)} \quad = \operatorname{tr} \left[M^{(2)}P_\lambda - M^{(1)}S_\lambda M^{(1)}P_\lambda \right],$$

where P_λ is the projection onto the eigenspace of λ, while S_λ is the *reduced resolvent* of M at λ, which is the holomorphic part of the resolvent of M at λ, defined by the properties

$$(48) \qquad S_\lambda P_\lambda = P_\lambda S_\lambda = 0; \qquad (M - \lambda\mathbf{I})S_\lambda = S_\lambda(M - \lambda\mathbf{I}) = \mathbf{I} - P_\lambda,$$

(in other words, S_λ is the inverse of $M - \lambda\mathbf{I}$ restricted to the orthogonal complement of the eigenspace of λ), and thus

$$(49) \qquad MS_\lambda = \mathbf{I} - P_\lambda + \lambda S_\lambda.$$

Now we will specialize a bit:

Assumption 1. The eigenvalue λ is such that the constant vector $\mathbf{1}$ spans the eigenspace of λ.

In this case, $P_\lambda = J_k/k$, where we recall that J_k is the $k \times k$ matrix of all 1s.
In addition,

Assumption 2. We will assume that $M(x) = D(x)M$, where $D(x)$ is an analytically varying diagonal matrix, $D(x) = D + D^{(1)}x + D^{(2)}x^2 + \dots$, where we say that the diagonal elements of $D^{(l)}$ are $\mathbf{d}^{(1)} = (d_1^{(l)}, \dots, d_k^{(l)})$.

Lemma 10.1. *Let $A = (A_{ij})$ be an $n \times n$ matrix. Then*

$$\operatorname{tr} AJ_n = \sum_{1 \le i,j \le n} A_{ij}.$$

Lemma 10.2. *Let $A = (A_{ij})$ be an $n \times n$ matrix, and let X be an $n \times n$ diagonal matrix. Then*

$$(XA)_{ij} \quad = A_{ij}X_{ii},$$
$$(XAX)_{ij} \quad = A_{ij}X_{ii}X_{jj}.$$

Lemma 10.3. *Let D be a diagonal matrix, with diagonal elements $d_1, \dots, d_n$. Then*

$$v^t Dv = \sum_{i=1}^n d_i v_i^2.$$

The proofs of the above lemmas are immediate.

Lemma 10.4. *Let P_v is the projection operator on the subspace generated by v (a unit vector). Then*

$$\operatorname{tr} M P_v = v^t M v.$$

In particular, if v is an eigenvector of M with eigenvalue λ, then $\operatorname{tr} M P_v = \lambda \|v\|$.

Proof. This follows by a direct computation, since when v is a unit vector, $(P_v)_{ij} = v_i v_j$. ◻

Lemma 10.5. *If v is an eigenvector of M with eigenvalue λ, then $M P_v = \lambda P_v$*

Lemma 10.6. *Suppose that λ has multiplicity 1, and $v(\lambda)$ is a unit vector generating the eigenspace of λ, and $M(t) = D(t)M$, where $D(t)$ is a diagonal matrix. Then*

$$\lambda'(M) = \lambda v^t(\lambda) D' v.$$

Proof. By Formula (46), we have

$$\lambda'(M) = \operatorname{tr} M' P_{v(\lambda)} = v^t(\lambda) M' v(\lambda) = \lambda v^t(\lambda) D' v.$$

◻

Corollary 10.7. *In the case when $v(\lambda) = \mathbf{1}$, we have:*

$$(50) \qquad \lambda^{(1)} = \frac{\lambda}{k} \sum_{j=1}^{k} d_j^{(1)}.$$

To compute the second derivative of λ, we use the formula (47) (we are assuming that λ is an isolated eigenvalue with eigenvector $v(\lambda)$, and $M(t) = D(t)M$, as before):

$$
\begin{aligned}
\lambda'' \quad &= \operatorname{tr}\left[M'' P_{v(\lambda)} - M' S_\lambda M' P_\lambda \right] \\
&= \lambda v^t D'' v - \operatorname{tr}\left[M' S_\lambda M' P_\lambda \right] \\
&= \lambda v^t D'' v - \lambda \operatorname{tr}\left[D' M S_\lambda D' P_\lambda \right] \\
&= \lambda v^t \left[D'' - D' M S_\lambda D' \right] v.
\end{aligned}
$$

We can now use the formula (49) to get:

$$(51) \qquad \lambda'' = \lambda v^t \left[D'' - D'(\mathbf{I} - P_\lambda) D' - \lambda D' S_\lambda D' \right] v.$$

In the special case where the eigenvector v is proportional to $\mathbf{1}$, we can rewrite the formula in coordinates in a simple way. To wit, any diagonal matrix D can be written (uniquely) as $D_0 + d\mathbf{I}$, where D_0 is such that $\operatorname{tr} D_0 = 0$. A simple computation then shows that

$$(52) \qquad \lambda'' = \frac{\lambda}{k} \left[\sum_{j=1}^{n} d_j'' - \sum_{j=1}^{n} (d_0')^2 - \lambda \mathbf{d}'^t S_\lambda \mathbf{d}' \right].$$

The case we are interested in is still more special, and that is where

Assumption 3.

$$D(x) = \begin{pmatrix} \exp(if_1 x) & & & \\ & \exp(if_2 x) & & \\ & & \ddots & \\ & & & \exp(if_k x) \end{pmatrix}.$$

Here, $\mathbf{d}^{(1)} = (if_1, if_2, \ldots, if_k)$, while $\mathbf{d}^{(2)} = -\frac{1}{2}(f_1^2, f_2^2, \ldots, f_k^2)$, and so, letting $\mathbf{f} = (f_1, \ldots, f_k)$,

$$(53) \qquad \lambda^{(2)} = \frac{\lambda}{k} \left[-\frac{1}{2} \|\mathbf{f}\|^2 + \|\mathbf{f}_0\|^2 + \lambda \mathbf{f}^t S_\lambda \mathbf{f} \right],$$

where, as before, $\mathbf{f}_0$ is the component of $\mathbf{f}$ orthogonal to constants.

To show our final estimates we shall need

Assumption 4. The matrix M is $\lambda > 0$ times a doubly stochastic matrix (this implies that the operator norm and the spectral radius of M are both equal to λ).

Theorem 10.8. *With assumptions as above, and, in addition, $\mathbf{f} = \mathbf{f}_0$ (that is $\sum_{j=1}^{k} f_j = 0$), then $\lambda^{(2)}$ is nonpositive.*

Proof. Since $\mathbf{f} = \mathbf{f}_0$, Equation (53) can be rewritten as

$$(54) \qquad \lambda^{(2)} = -\frac{\lambda}{2k} \left[-\|\mathbf{f}_0\|^2 - 2\lambda \mathbf{f}_0^\perp S_\lambda \mathbf{f}_0 \right] = -\frac{\lambda}{2k} \left[\mathbf{f}_0^t \left(-\mathbf{I} - 2\lambda S_\lambda \right) \mathbf{f}_0 \right].$$

If we regard S_λ as an operator on the orthogonal complement to $\mathbf{1}$, then by equations (48) and (49), $S_\lambda(\lambda \mathbf{I}_0 - M_0) = -\mathbf{I}_0$. Let $v = -S_\lambda \mathbf{f}_0$. Then the term in square brackets in Eq. 54 can be rewritten as:

$$(55) \qquad v^t (\lambda \mathbf{I}_0 - M_0)^t \left(-\mathbf{I} - 2\lambda S_\lambda \right) (\lambda \mathbf{I}_0 - M_0) v = v_t \left(\lambda^2 \mathbf{I}_0 - M_0^t M_0 \right) v,$$

where we have used the fact that for any matrix A and any vector v, $v^t A v = v^t A^t v$. The quadratic form $\lambda^2 \mathbf{I}_0 - M_0^t M_0$ is positive semi-definite, since the biggest eigenvalue of the symmetric matrix $M_0^t M_0$ is equal to the square of the operator norm of M_0, which, in turn, is no greater then λ, by Assumption 4 (since $M^t M$ is λ^2 times a doubly stochastic matrix). □

Remark 10.9. *In the statement of Theorem 10.8, the word "non-positive" can be improved to "negative" under the further assumption that M is irreducible, primitive, and normal.*

Proof. Since the orthogonal complement to the subspace generated by the vector $\mathbf{1}$ is invariant under M, it follows that M_0 is also normal, and so its operator norm is equal to its spectral radius μ. Under the assuptions of irreducibility and primitivity, Perron-Frobenius theory tells us that $|\mu| < \lambda$. □

11. TOPOLOGICAL ENTROPY

Consider a graph G, and consider a positive function f on its vertices. For each cycle c we let $F(c)$ to be the sum of values of f over c, and we want to know how many c are there for which $F(c) \leq L$. We denote that number by $N(f, L)$, and we ask ourselves how $N(f, L)$ behaves asymptotically as L tends to infinity. To understand $N(L, f)$, we consider first the matrix $U(f) = D(u^{f_1}, \ldots, u^{f_n}) A(G)$. As

before, we observe that the coefficient of u^r in $\operatorname{tr} U^n(f)$ is the number of cycles of (combinatorial) length n, for which $F(c) = r$. Write a formal series

$$L(f, u) = \sum_n \operatorname{tr} U^n(f).$$

This series converges for sufficiently small u, and can there be written in closed form as $L(f, u) = \operatorname{tr} (\mathbf{I} - U(f))^{-1}$, from which it follows that the exponential rate f growth of $N(c)$ is equal to negative logarithm of the radius of convergence of $L(f, u)$ – we call this the *entropy of G, f* – which, in term, is equal to the smallest positive real value of u, such that the spectral radius of $U(f)$ is equal to 1. Since it is more convenient to deal with analytic functions (which $L(f, u)$ is not, for arbitrary real values of f_i, so we write $u = \exp -s$, and now ask for the abscissa of convergence of $L(f, \exp -s)$. This will give us the entropy. In this section we use perturbation methods in a rather straightforward way to get explicit information on the entropy.

Let A be an $n \times n$ non-negative primitive irreducible matrix. Let $f_1, \ldots, f_n$ be a collection of weights. We then define the matrix $E(s, \mathbf{f})$ to be the diagonal matrix whose ii-th element is equal to $\exp(-sf_i)$. Define $M(s, \mathbf{f})$ to be $M(s, \mathbf{f}) = E(s, \mathbf{f})A$. We are interested in $\rho(s, \mathbf{f})$: the spectral radius of $M(s, \mathbf{f})$. By Perron-Frobenius theory we know that there is a real eigenvalue of $M(s, \mathbf{s})$ equal to $\rho(s, \mathbf{f})$, and the eigenvector v_ρ of this eigenvalue is positive.

Lemma 11.1.

$$(56) \qquad \frac{\partial \rho}{\partial s} = -\rho v_\rho^t D(f_1, \ldots, f_n)v.$$

For positive $\mathbf{f}$, $\frac{\partial \rho}{\partial s} < 0$.

Proof. This follows immediately from Lemma 10.4 and the positivity of ρ and v_ρ. $\qquad\square$

Lemma 11.2. *We have the following expression for the gradient of ρ with respect to $\mathbf{f}$:*

$$(57) \qquad \nabla_{\mathbf{f}} \rho = -s\rho(v_1^2, \ldots, v_n^2),$$

where $v_\rho = (v_1, \ldots, v_n)$.

Proof. We note that

$$\frac{\partial M}{\partial f_i} = -sD(0, \ldots, 1, \ldots, 0)M,$$

where the 1 is in the i-th place. Thus, by formula (46) we have

$$\frac{\partial \rho}{\partial f_i} = -sv_\rho^t D(0, \ldots, 1, \ldots, 0)Mv = -s\rho v_i^2.$$

$\qquad\square$

This can be restated as saying that the derivative of ρ in the direction of a vector $\mathbf{g}$ is equal to $-\rho s v_\rho^t D(\mathbf{g})v$.

This gives us the following important corollary:

Corollary 11.3. *Consider deformations g keeping the sum of f_i fixed. Then the critical points of ρ occur precisely for those ρ for which $|v_i| = |v_j|$, for any i, j.*

We can also compute the second directional derivative of ρ. Indeed, let $\mathbf{g} = (g_1, \ldots, g_n)$ be the direction vector, so that we want to compute the second derivative with respect to t of $\rho(s, \mathbf{f} + t\mathbf{g})$ at $t = 0$. To do this, we use the formula (47):

$$(58) \qquad \rho'' = \operatorname{tr} \left[M'' P_{v(\rho)} - M' S_\rho M' P_\rho \right].$$

Note that (as in the proof of Lemma 11.2)

$$(59) \qquad M' = -sD(g_1, \ldots, g_n)M,$$

while

$$M'' = s^2 D(g_1^2, \ldots, g_n^2)M,$$

and so

(60)
$$\operatorname{tr} M'' P_{v(\rho)} = s^2 \rho v^t D(g_1^2, \ldots, g_n^2)v = s^2 \rho \left\{ D(g_1, \ldots, g_n)v \right\}^t \left\{ D(g_1, \ldots, g_n)v \right\}.$$

To understand the second term of the right-hand side of Eq. (58), first note that (by Eq. (59))

$$M' S_\rho M' P_\rho = D(g_1, \ldots, g_n)M P_\rho = \rho s^2 D(g_1, \ldots, g_n)M S_\rho D(g_1, \ldots, g_n)P_\rho,$$

, where the second equality is by Lemma 10.5. Now

$$(61) \qquad \operatorname{tr} M' S_\rho M' P_\rho \quad = \rho s^2 v(\rho)^t D(g_1, \ldots, g_n)M S_\rho D(g_1, \ldots, g_n)v$$

$$(62) \qquad\qquad\qquad = \rho s^2 \left\{ D(g_1, \ldots, g_n)v \right\}^t M S_\rho \left\{ D(g_1, \ldots, g_n)v \right\}.$$

Putting together Eq. (60) and Eq. (61), we see that

$$(63) \qquad \rho'' = \rho s^2 \left\{ D(g_1, \ldots, g_n)v \right\}^t (\mathbf{I} - M S_\rho) \left\{ D(g_1, \ldots, g_n)v \right\}$$

Using the formula (49) equation (63) simplifies further to:

$$(64) \qquad \rho'' = \rho s^2 \left\{ D(g_1, \ldots, g_n)v \right\}^t \left(P_{v(\rho)} - \rho S_\rho \right) \left\{ D(g_1, \ldots, g_n)v \right\}$$

The following lemma is not surprising:

Lemma 11.4. *The quadratic form given $P_v - \rho S_\rho$ is positive-definite.*

Proof. On the span of v, the projection operator P_v is equal to the identity, whilst the reduced resolvent S_ρ vanishes. On the orthogonal complement, the projection operator vanishes, so since the Perron-Frobenius eigenvalue ρ is positive, we need to show that S_ρ is negative definite. Consider a vector w, in the orthogonal complement of v. Such a w is equal to $(\rho\mathbf{I} - M)z$, for some z orthogonal to v. So,

$$w^t S_\rho w = z^t (\rho I - M)z,$$

So, it will suffice to show that $(\rho I - M)$ is negative-definite. Suppose not. Then there exists a z_0, such that $z_0^t M z_0 \geq \rho \|z_0\|^2$. By the argument in the proof of theorem 10.8, we see that $\|M z_0\| \leq \rho \|z_0\|$. So, $z_0^t M z_0 \geq \rho \|z_0\|^2$ implies that $\langle z_0, M z_0 \rangle \geq \rho \|z_0\|^2$, and hence that z_0 is an eigenvactor of M with eigenvalue ρ, which is impossible by assumtion that M is irreducible and primitive. $\qquad\square$

We finish with

Theorem 11.5. *Let $s_0(\mathbf{f})$ be the unique s such that $\rho(s_0, \mathbf{f})$ is equal to 1. Then s_0 is a convex function of $\mathbf{f}$, and hence assumes a unique minimum on each linear subspace of values of $\mathbf{f}$. In particular, if we restrict to the the subspace F_0, where the sum of the values of of $\mathbf{f}$ is equal to 1, then the minimum is achieved at the point where*

$$(65) \qquad f_i = \frac{\log(A\mathbf{1})_i}{\sum_i \log(A\mathbf{1})_i},$$

in which case the entropy is equal to $\sum \log(A\mathbf{1})_i$.

Proof. The convexity of s_0 follows from Lemma 11.4 and Lemma 11.1. The point at which the minimum is achieved is computed easily using Corollary 11.3, as is the value of entropy. $\square$

12. Applications to Groups and other objects

The asymptotic results in the previous sections apply directly to the question of the growth of homology classes in the free groups, and give in some sense complete information:

Observation 12.1. *We see that the asymptotic order of growth of any two fixed homology classes is the same.*

Observation 12.2. *Theorem 7.1 shows in particular that a random long cycle is equidistributed among the vertices of a regular graph.*

Observation 12.3. *We see that the order of growth the number of words length n in any fixed homology class in F_k is asymptotic to $c_k(2k-1)^n/n^{k/2}$, where c_k is easily computed using the expression for σ in the statement of Theorem 5.1, keeping in mind that*

$$c_{F_k} = \frac{k}{\sqrt{2k-1}},$$

where c is the parameter in the statements of theorems of the last two sections. Alternately, Theorem 7.1 can be used.
(c) We can compute other growth functions. For example, let $h : F_n \to \mathbf{Z}$ be the "total exponent" homomorphism, i.e. if $F_n =< a_1, \ldots, a_n >$, then $h(a_i) = 1$. We see that the generating function for the preimages of $j \in Z$ is given by

$$\left(2\sqrt{2n-1}\right)^k R_k\left(\frac{n}{\sqrt{2n-1}}; x, \ldots, x\right) = \left(2\sqrt{2n-1}\right)^k R_k\left(\frac{n}{\sqrt{2n-1}}; x\right).$$

Observation 12.4. *Instead of cyclically reduced words, it is perhaps more natural to study conjugacy classes (ordered by their cyclically reduced length). It seems futile to seek any enumeration as neat as Theorem 2.3, however, since the relationship between the number $\mathcal{C}_k$ of conjugacy classes of words of length k and the number of cyclically reduced words $\mathcal{W}_k$ is:*

$$(66) \qquad \mathcal{C}_k = \frac{\mathcal{W}_k}{k} + O(\sqrt{\mathcal{W}_k}),$$

it is clear that the asymptotic results are the same for the two problems. For more on this subject, see Section 13 and the sequel.

Observation 12.5. *Counting conjugacy classes is a problem closely related to that of counting closed geodesics on manifold. In the context of compact hyperbolic surfaces, it was observed by P. Sarnak (see, for example, [RS94]) that among all geodesics shorter than L, null-homologous geodesics are more numerous than those in any other prescribed homology class (that is, while the ratio of the two quantities approaches* 1, *the difference is asymptotically positive). The results of the current note provide a certain justification for this, since any limiting distribution likely to arise in this context is, for reasons of symmetry, likely to be unimodal, with the mode at* **0**. *Certainly this is true of the normal distribution, though even in this case, a careful analysis of the error terms is required.*

13. Counting conjugacy classes

Consider a finitely presented group G. Let g be an element of G. We define the *reduced length* of g – denoted by $|g|$ – to be the length of the shortest word in the generators of G representing g. We define the *length up to conjugacy* of g – denoted by $|g|_c$ – to be the minimum of $|h|$, the minimum being taken over all group elements h conjugate to g. Length up to conjugacy is obviously invariant under conjugation, and we will also use the term to apply to conjugacy classes.

$$\mathcal{N}_G(r) = \left|\{g \in G \mid \quad |g| = r\}\right|,$$
$$\mathcal{C}_G(r) = \left|\{g \in \mathcal{N}_G(r) \mid |g|_c = r\}\right|,$$
$$\mathcal{CC}_G(r) = \left|\{C \in G/\text{conjugacy} \mid |C|_c = r\}\right|.$$

The subscript G will be omitted whenever the group G is obvious from context.

Given a sequence $A = a_0, \dots, a_i, \dots$, we can define a *generating function* $\mathcal{F}[A]$, by

$$\mathcal{F}[A](z) = \sum_{i=0}^{\infty} a_i z^i.$$

There is frequently confusion as to whether the generating function is a holomorphic function or an element of the ring of formal power series. In this section "generating function" will mean a function analytic at $0 \in \mathbf{C}$.

The three counting functions above give rise to corresponding generating functions $\mathcal{F}[\mathcal{N}_G]$, $\mathcal{F}[\mathcal{C}_G]$, $\mathcal{F}[\mathcal{CC}_G]$. Our real interest will lie in the last of these; the first one has been the most extensively studied, and the result most relevant to us is:

Fact 1. If G is an *automatic* group, then the generating function $\mathcal{F}[\mathcal{N}_G]$ is a rational function.

For definitions and properties of automatic groups, see [ECHLPT92].

Fact 2. (Gromov, Epstein) If G is an automatic group, then the generating function $\mathcal{F}[\mathcal{C}_G]$ is a rational function.

Facts 1 and 2 might lead us to expect that $\mathcal{F}[\mathcal{CC}_G]$ is, likewise, rational, but in fact the opposite seems to be the case, and we are led to:

Conjecture 13.1. *Let G be a word-hyperbolic group. The $\mathcal{F}[\mathcal{CC}_G]$ is rational if and only if G is virtually cyclic (*elementary *in the terminology of* [Gromov87]*).*

In the sequel, this conjecture is supported by the complete analysis of the case where G is F_k – the free group on k generators.

14. GROWTH FUNCTIONS FOR FREE GROUPS

Let F_k be the free group on k generators. The following is obvious:

Fact 3. $\mathcal{N}_{F_k}(r) = 2k(2k-1)^{r-1}$.

Theorem 1.1 says that

$$\mathcal{C}_{F_k}(r) = (2k-1)^r + 1 + (k-1)[1 + (-1)^r].$$

Corollary 14.1.

$$\mathcal{F}[\mathcal{C}_{F_k}](z) = \frac{1}{1-(2k-1)z} + \frac{1}{1-z} + \frac{2(k-1)}{1-z^2} - 2k.$$

In order to compute $\mathcal{CC}_{F_k}(r)$ it is enough to notice the following:

Theorem 14.2.

$$r\mathcal{CC}(r) = \sum_{d \mid r} \phi(d)\mathcal{C}(r/d),$$

where ϕ denotes the Euler totient function.

Proof. The theorem is a trivial consequence of Burnside's lemma, stated below as Theorem 14.3 for convenience, applied to the action of the cyclic group $\mathbf{Z}/(r\mathbf{Z})$ on the set of cyclically reduced words of length r. $\square$

Theorem 14.3. *Let G be a finite group acting on a finite set X. For $g \in G$ let $\psi(g)$ denote the number of $x \in X$, such that $g(x) = x$. Then the number of orbits of X under the G-action is*

$$\frac{1}{|G|} \sum_{g \in G} \psi(g).$$

We now have the following general observation:

Theorem 14.4. *Suppose we have three sequences $A = \{a_i\}$, $B = \{b_j\}$, and $C = \{c_k\}$, satisfying*

$$a_n = \sum_{d \mid n} c_d b_{\frac{n}{d}}.$$

Then

$$\mathcal{F}[A](z) = \sum_{d=1}^{\infty} c_d \mathcal{F}[B](x^d).$$

Proof. On the level of formal power series, the statement is clear by expanding the left hand side. Otherwise, if the radius of convergence of $\mathcal{F}[A]$ is r_a, then the radius of convergence of $G_d[A]$, defined as $G_d[A](z) = \mathcal{F}[A](z^d)$ is, by Hadamard's criterion, equal to $r_a^{1/d}$, so all of $G_d[A]$ converge on the disk of radius $R_a = \min(r_a, 1)$ around the origin. Since the series on the right hand side converges at 0 (since all the terms vanish), it converges uniformly on compact subsets of the disk of radius R_a around the origin. $\square$

Corollary 14.5. *Let $\mathcal{H}$ be the generating function of the sequence $h_r = r\mathcal{CC}(r)$. Then*

$$\mathcal{H}(z) = 1 + \sum_{d=1}^{\infty} \phi(d)\mathcal{F}[\mathcal{C}](z^d).$$

We can combine all of the above results into the following conclusion:

Theorem 14.6. *The generating function $\mathcal{H}$ as in the statement of corollary 14.5 can be expanded as:*

$$\mathcal{H} = 1 + (k-1)\frac{x^2}{(1-x^2)^2} + \sum_{d=1}^{\infty} \phi(d)\left(\frac{1}{1-(2k-1)x^d} - 1\right).$$

In particular, $\mathcal{H}$ has an infinite number of poles, and is not a rational function. The generating function $\mathcal{F}[\mathcal{CC}_{F_k}]$ can be written as

$$\mathcal{F}[\mathcal{CC}_{F_k}](z) = \int_0^z \frac{\mathcal{H}(t)}{t}dt,$$

and so is not a rational function either.

Proof. The expression for $\mathcal{H}$ is fairly obvious, with the comment that the second summand is a consequence of the fact that

$$\sum_{d \mid n} \phi(d) = n.$$

That $\mathcal{H}$ has an infinite number of poles follows from the observation that the d-th term in the third summand has its d poles on the circle $|z| = (2k-1)^{-1/d}$, while the first two summands are analytic in the open unit disk. The expression for $\mathcal{F}[\mathcal{CC}_{F_k}]$ is immediate. $\square$

Remark. Various people, when shown Theorem 14.6, appeared to believe that it contradicts [Gromov87, Theorem 5.2D]. In fact (as pointed out by Greg McShane), Gromov's function $[N]_k$ is *not* (as the common misunderstanding has it) the same as $\mathcal{CC}_G(r)$ in the case of a free group, but *is* the same as $\mathcal{C}_G(r)$.

14.1. **Some further comments.** The following observation is quite obvious:

Observation 14.7. *Let G_1 and G_2 be two groups. Then,*

$$\mathcal{F}[\mathcal{CC}_{G_1 \times G_2}](z) = \mathcal{F}[\mathcal{CC}_{G_1}](z)\mathcal{F}[\mathcal{CC}_{G_2}](z).$$

Observation 14.7 has some consequences:

Theorem 14.8. *Let G_1 and G_2 be two groups, then if $\mathcal{F}[\mathcal{CC}_{G_1}]$ is rational, while $\mathcal{F}[\mathcal{CC}_{G_2}]$ is not, then $\mathcal{F}[\mathcal{CC}_{G_1 \times G_2}]$ is not rational. If both $\mathcal{F}[\mathcal{CC}_{G_1}]$ and $\mathcal{F}[\mathcal{CC}_{G_2}]$ are rational, then so is $\mathcal{F}[\mathcal{CC}_{G_1 \times G_2}]$.*

Corollary 14.9. *If $G_1 = \mathbf{Z}^n$ and G_2 is a finite group, then $\mathcal{F}[\mathcal{CC}_{G_1 \times G_2}]$ is rational.*

Remark. It is not clear whether $\mathcal{F}[\mathcal{CC}_G]$ is rational when G is a Bieberbach group – most likely this depends on the choice of the generating set, as conjectured by D. B. A. Epstein.

Corollary 14.10. *If $G_1 = F_k$ and G_2 is a direct product of finite groups and infinite cyclic groups, then $\mathcal{F}[\mathcal{CC}_{G_1 \times G_2}]$ is irrational.*

Theorem 14.11. *If $G = F_{k_1} \times F_{k_2} \times \ldots \times F_{k_n}$, then $\mathcal{F}[\mathcal{CC}_G]$ is irrational (with respect to the "obvious" generating set).*

Proof. This is an immediate consequence of Theorem 14.6. $\qquad\square$

15. Primitive conjugacy class zeta function

One can compute a zeta-function analogous to that of Ihara for the numbers of *primitive* conjugacy classes of a given length (a primitive class is one which is not the power of a smaller class), using, essentially, the elementary method described by Stark and Terras, [ST96], as applied to the graph constructed in Section 1. This function turns out to be rational (in fact, there is a simple formula for it, see Theorem 15.1). More precisely, consider

$$(67) \qquad \zeta(G)^{-1} = \prod_{[c]}(1 + u^l(c)),$$

where $[c]$ denotes the equivalences classes of primitive cycles, where two cycles are considered equivalent if one can be obtained from the other by a rotation.

A computation then shows that

$$(68) \qquad \zeta(F_r) = (1 - u^2)^{r-1}(1 - u)(1 - (2r - 1)u).$$

The computation goes as follows:

First, note that

$$(69) \qquad \log \zeta(G) = \sum_{[c]} \sum_{i=1}^{\infty} \frac{1}{i} u^{il(c)},$$

and thus

$$(70) \qquad u\frac{d \log \zeta(G)}{du} = \sum_{[c]} \sum_{i=1}^{\infty} l(c) u^{il(c)}.$$

The above can be rewritten (note that the sum is now over primtive cycles, and not equivalence classes thereof):

$$(71) \qquad u\frac{d \log \zeta(G)}{du} = \sum_{c} \sum_{i=1}^{\infty} u^{il(c)}.$$

But note that the right hand side is simply the ordinary generating function for *all* cycles:

$$(72) \qquad u\frac{d \log \zeta(G)}{du} = \sum_{i=1}^{\infty} N_i u^i,$$

where N_i is the number of cycles of length i in G, and this generating function was computed in Section 1:

$$(73) \qquad \sum_{i=i}^{\infty} N_i u^i = \frac{1}{1 + (2r - 1)u} + \frac{r}{1 - u} + \frac{r - 1}{1 + u}.$$

The formula 68 now follows by a straightforward integration.

An quick examination of the above argument shows that the formula 68 is a special case of the following result:

Theorem 15.1. *Let G be a finite graph, and let ζ_G be the zeta function defined by formula 67. Let $A(G)$ be the adjacency matrix of G. Then*

$$(74) \qquad\qquad \zeta_G(u) = \det\left(I - uA(G)\right).$$

In other words, the zeta function is essentially the characteristic polynomial of $A(G)$.

Proof. The argument above up to Equation 72 is completely general. On the other hand, the right hand side of equation 72 can be rewritten as:

$$
\begin{aligned}
\sum_{i=1}^{\infty} N_i u^i \qquad &= \textstyle\sum_{i=1}^{\infty} \operatorname{tr} A(G)^i u^i \\
&= \operatorname{tr}\left[-I + \textstyle\sum_{i=0}^{\infty} \left[A(G)u\right]^i\right] \\
&= \operatorname{tr}\left[-I + (I - uA(G))^{-1}\right] \\
&= \operatorname{tr}\left(uA(G)(I - uA(G))^{-1}\right).
\end{aligned}
$$

Thus,

$$\frac{d\log\zeta(G)}{du} = \operatorname{tr}\left(A(G)(I - uA(G))^{-1}\right),$$

, and so it follows that

$$\zeta(G) = C\det(I - uA(G)),$$

where C is a constant of integration, seen to be equal to 1 by computing both sides at $u = 0$. □

References

AHU75.   A. V. Aho, J. E. Hopcroft, J. D. Ullman. *Design and analysis of computer algorithms*, Addison-Wesley series in Computer Science and Information Processing. Addison-Wesley Publishing Company, Reading, Massachussetts, 1975.

BP97.   Martine Babillot and Marc Peigné. *Closed geodesics on homology classes on hyperbolic manifolds with cusps*, Comptes Rendus Acad. Sci. Paris, Sér. I Math. **324** 1997, no. 8, pp. 901-906.

Bass92.   Hyman Bass. *On the Ihara-Selberg zeta function of a tree lattice*, International Journal of Mathematics, **3**(1992), pp. 717-797.

BruRy91.   R. Brualdi and H. J. Ryser. *Combinatorial Matrix Theory*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, 1991.

Buser92.   Peter Buser. *Geometry and spectra of compact Riemann surfaces*, Progress in Mathematics 106. Birkhäuser Boston, Inc., Boston, MA 1992.

ECHLPT92.   D. B. A. Epstein, J. W. Cannon, D. F. Holt, S. V. F. Levy, M. S. Paterson, W. P. Thurston. *Word Processing in groups*, Jones and Bartlett Publishers, Boston, Massachussetts, 1992.

FellerII.   William Feller. *An Introduction to Probability Theory and its Applications, Volume II*, second edition, John Wiley, New York, 1971.

FH91.   William Fulton and Joe Harris. *Representation Theory*, Springer-Verlag, New York-Berlin, 1991.

Gromov87.   M. Gromov. *Hyperbolic groups*, in Essays in group theory (S. M. Gersten, editor), MSRI publications **8**, Springer, New York-Berlin, 1987.

Ihara66.   Y. Ihara. *On discrete subgroups of the two by two projective linear group over* p-*adic fields* Journal of the Mathematical Society of Japan, **18**(1966), pp. 219-235.

Kato66.   Tosio Kato. *Perturbation Theory for Linear Operators*, Die Grundlehren der mathematischen Wissenschaften **132**, Springer-Verlag, 1966.

PS87.   Ralph Phillips and Peter Sarnak. *Geodesics in homology classes*, Duke Math. Journal **55** (1987), no. 2, pp. 287-297.

Rivlin90.   Theodore J. Rivlin. *Chebyshev Polynomials: From Approximation Theory to Algebra and Number Theory*, second edition, Wiley Interscience, 1990.

RS94.   Michael Rubinstein and Peter Sarnak. *Chebyshev's bias*, Experiment. Math. **3** (1994), no. 3, pp. 173-197

Schur73.   Issai Schur. *Arithmetischen über die Tschebyscheffschen Polynome*, in *Gesammelte Abhandlungen*, **III**, Springer, Berlin, 1973, pp 422-453.

Serre89.   Jean-Pierre Serre. *Abelian ℓ-adic representations and elliptic curves*, Addison-Wesley Advanced Book Classics, 1989.

Shiryaev96.   A. N. Shiryaev. *Probability*, second edition, Springer, New York, 1996.

ST96.   Harold Stark and Audrey Terras. *Zeta Functions of finite graphs and coverings*, Adavnces in Mathematics **121** (1993).

Vardi93.   Ilan Vardi. *Dedekind Sums have a limiting distribution*, International Mathematics Research Notices 1993, no. 1, pp. 1-12.

Weil51.   André Weil. *L'intégration dans les groupes topologiques et ses applications*, 2ème édition, Hermann, Paris, 1953.

MATHEMATICS INSTITUTE, WARWICK UNIVERSITY, COVENTRY, THE UNITED KINGDOM

Current address: Mathematics Department, University of Manchester, Manchester, The United Kingdom

E-mail address: `igor@maths.warwick.ac.uk`